

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/09, 15/12, 15/33, 15/64		A1	(11) International Publication Number: WO 96/09378 (43) International Publication Date: 28 March 1996 (28.03.96)
(21) International Application Number: PCT/US95/11511		(81) Designated States: AU, BG, BR, BY, CA, CN, CZ, FI, HU, JP, KR, MX, NO, NZ, PL, RO, RU, SG, SI, UA, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(22) International Filing Date: 8 September 1995 (08.09.95)		Published <i>With international search report.</i>	
(30) Priority Data: 08/324,243 19 September 1994 (19.09.94) US			
(71) Applicant: THE GENERAL HOSPITAL CORPORATION [US/US]; 55 Fruit Street, Boston, MA 02114 (US).			
(72) Inventor: SEED, Brian; Apartment 5J, Nine Hawthorne Place, Boston, MA 02114 (US).			
(74) Agent: LECH, Karen, F.; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02114 (US).			
(54) Title: OVEREXPRESSION OF MAMMALIAN AND VIRAL PROTEINS			
(57) Abstract			
<p>The invention features a synthetic gene encoding a protein normally expressed in mammalian cells wherein at least one non-preferred or less preferred codon in the natural gene encoding the mammalian protein has been replaced by a preferred codon encoding the same amino acid.</p>			

- 55 -

1. A synthetic gene encoding a protein normally expressed in mammalian cells wherein at least one non-preferred or less preferred codon in the natural gene encoding said mammalian protein has been replaced by a preferred codon encoding the same amino acid.
5
2. The synthetic gene of claim 1 wherein said synthetic gene is capable of expressing said mammalian protein at a level which is at least 110% of that expressed by said natural gene in an in vitro mammalian cell culture system under identical conditions.
10
3. The synthetic gene of claim 1 wherein said synthetic gene is capable of expressing said mammalian protein at a level which is at least 150% of that expressed by said natural gene in an in vitro cell culture system under identical conditions.
15
4. The synthetic gene of claim 1 wherein said synthetic gene is capable of expressing said mammalian protein at a level which is at least 200% of that expressed by said natural gene in an in vitro cell culture system under identical conditions.
20
5. The synthetic gene of claim 1 wherein said synthetic gene is capable of expressing said mammalian protein at a level which is at least 500% of that expressed by said natural gene in an in vitro cell culture system under identical conditions.
25
6. The synthetic gene of claim 1 wherein said synthetic gene is capable of expressing said mammalian protein at a level which is at least ten times that expressed by said natural gene in an in vitro cell culture system under identical conditions.
30

- 56 -

7. The synthetic gene of claim 1 wherein at least 10% of the codons in said natural gene are non-preferred codons.

8. The synthetic gene of claim 1 wherein at least 5 50% of the codons in said natural gene are non-preferred codons.

9. The synthetic gene of claim 1 wherein at least 50% of the non-preferred codons and less preferred codons present in said natural gene have been replaced by 10 preferred codons.

10. The synthetic gene of claim 1 wherein at least 90% of the non-preferred codons and less preferred codons present in said natural gene have been replaced by preferred codons.

15 11. The synthetic gene of claim 1 wherein said protein is a retroviral or lentiviral protein.

12. The synthetic gene of claim 11 wherein said protein is an HIV protein.

20 13. The synthetic gene of claim 12 wherein said protein is selected from the group consisting of gag, pol, and env.

14. The synthetic gene of claim 13 wherein said protein is gp120 or gp160.

25 15. The synthetic gene of claim 1 wherein said protein is a human protein.

- 57 -

16. A method for preparing a synthetic gene encoding a protein normally expressed by mammalian cells, comprising identifying non-preferred and less-preferred codons in the natural gene encoding said protein and
5 replacing one or more of said non-preferred and less-preferred codons with a preferred codon encoding the same amino acid as the replaced codon.

Syngp120mn

1 CTCGAGATCC ATTGTGCTCT AAAGGAGATA CCCCCCAGA CACCCCTCAC
 51 TCGGGTGCCT AGCTGCCCAAG CCTGAGGCAA GAGAAGGCAA GAAACCATGC
 101 CCATGGGTC TGTGAAACCT CTGGCCACCT TGTACCTGCT CGGAGATGCG
 151 CTGCTTCCG TGTAGCCAC CGAGAAGCTG TGGGTGACCG TGTACTACGG
 201 CGTGCCCCCTG TGTAGGAGG CGACCACAC ACCTGTTGAA CGAGGGACG
 251 CCAAGGGCTA CGACACCCAG CTGACACACG TGTGGCCAC CGAGGGCTGC
 301 CTGCCCCACG ACGGAAACCC CGAGGAGCTG GAGCTGTGA ACGTGACCSA
 351 GAACTTCAAC ATGTGGAAAGA ACAACATGCT CGAGCAGATG CATGAGGACA
 401 TCATCACCT GTGGGACCCAG AGCTGAAAGC CCTGCTGTAA CCTGACCCCC
 451 CTGTGCTGA CCTGAAACTS CACCGACCTG AGGAACACCA CGAACACCAA
 501 CAACAGCACC GCACACACCA ACAGCACACG CGAGGGCACC ATCAAGGGCG
 551 CGGAGATGAA CAACTGCAGC TTCAACATCA CGACCAAGCAT CGCGGACAAG
 601 ATGCAGAAGG ATGACGCTT CCTGATCAAG CTGGATATCG TTAGCATCGA
 651 CAACGACAGC ACCACCTACC CCTGATCTC CTGCAACACCG ACCGTGATCA
 701 CCCAGGCTG CGCCAAAGATC AGCTTGAGC CGATCCCCAT CGACTACTGC
 751 CGCCCCCGC CCTGGCCAT CCTGAACTGAA AACGACAAAGA AGTTGAGGG
 801 CAAGGGCAGC TCAAGAAAGC TTAGGACCCGT CGACTGACCC CACGGCATCC
 851 CGCCGGTGTG TGGCAGCCAG CCTGGCTGA AGGGCAGGCT CGGGGAGGAG
 901 GAGGTGCTGA TCCCAGCCG ACAAACCCCA AGACCATCAT
 951 CCTGGACCTG AATGAGAGGG TGAGATCAA CTGCAACCGT CGCAACTACA
 1001 ACAACGSCAA CGCCATCCAC ATCGGCCCCCG CGCGGCGCTT CTACACCCACC
 1051 AAGAACATCA TCGGCACCAT CGCCGAGGCG CACTGCAACCA TCTCTAGAGC
 1101 CAAGTGGAAC GACACCTGC CGCAGATCTG GAGCAAGCTG AAGGAGGAGT
 1151 TCAAGAACAA GACCATCGTG TTCAACCCAGA CGAGGGCGG CGACCCCCAG
 1201 ATCGTGATGC ACAGCTTCAA CTGGGGCGGC GAATTCTT ACTGCAACAC
 1251 CGCCCCCGTGC TTCAACAGCA CCTGGAAACGG CAAACACCC TGGAAACAACA
 1301 CGACCGGGAG CAACACAAAT ATTACCTCC AGTGGAGAT CAAGCAGATC
 1351 ATCAACATGT CGCAGGAGGT CGCCAAAGCCG ATGTACGCCCC CGCCCATCGA
 1401 CGGCCAGATC CGGTGGAGCA CGAACATCAC CGGTGCTG CTGACCCCCGG
 1451 ACGGGCGSCAA CGACACCCGAC ACCAACGACA CGGAAATCTT CGGGGGGGC

2/12

1501 GCGGGCGACA TCGCCGACAA CTGGAGATCT GAGCTTACA AGTACAAAGGT
1551 CCTGACCGATC GAGCCCCCTGG CCTTGCCCCC CACCAAGCCC AAGCCCCCC
1601 TGGTCCAGCG CAGAAGGCCT TAAAGCCCCC CC (SEQ ID NO:34)

FIG 1
(SHEET 2 OF 4)

Syngp160mn

1 AUCGAGAAGC TGTGGGTGAC CTTTCTACTAC CGCGTACCGG TGTGGAAAGGA
 51 GCGCACCCACG AGCTTGTTCG GCGCCGACCGA CGCGAAGGGG TACGACACCS
 101 ATGTGCAACAA CTTGTGGGGC ACCGGAGGTTG GCGTGCCTCAC CGACCCCGAAC
 151 CGCGAGGAGG TTGAGCTTGTG AAACGTCACG GAGAACTTCA ACATGTGGAA
 201 GAACAAACATG CTGGAGGAGA TGTATGGAGA GATCATTCAGC CTGTGGGACC
 251 AGAGCTGAA GCGCTTGTG AGCTGACCGG CCTGTGCGT GACCTTCAC
 301 TCGACCGTGC TGGGGACACG GACCAACACG AACAAACAGCA CGCGCGGAA
 351 CAACGCGAAC AGCGAGGGCA CGATCAAGGG CGCGCGAGATG AAUAACGTCG
 401 GTTTCACAT GACCGACCGG ATCGCGGACA AGATCGAGAA CGAGTGACCC
 451 TTGTGTAACA AGCTGGATAT CCTGAGGATC GACAAACGAA CGACCGACG
 501 CGCGCTGATG TGTGCGAACG CGACCGTGT GAGGGAGGG CGCGCGGAAAGA
 551 TGTGCTTGGG GCGCTTGGG AGCGAGGAGT CGCGCGGGCG CGCGCTTGGC
 601 ATCGTGTAACT GCGACGACAA GAACTTCAAG GCGAAGGGG GCGTCAAGAA
 651 CCTGACCGGG ATCGAGGCG CGACCGGGAT CGCGCGCGT GCGAGGACCC
 701 AGCTCTGTG GAAAGGGGGG CGCGCGGGGG AGGAGGTGGT GATTCGGAGC
 751 GAGGACTTGA CGCGACCGG CGAGGACCGT ATCGTGTGACG TGAATGAGAO
 801 CGTGCAGATG AACCGACCGG CGCGCGACTA CGAAGAGGCG AGCGCGACCG
 851 ACATCGGGCG CGCGCGCGCG TTGTACACCA CGAAGAACAT CGCGCGACCG
 901 ATCGCGCGGG CGCGACCGAA GATGTGAGA GCGAAGTGGG AGCGACCGGT
 951 CGCGCGAGAT GCGAGGAGCA GTTGTGAGCA CGCGCGGGGG AGCGACCGT
 1001 TGTGCAACTA GAGGAGGGG CGCGCGGGGG AGATCTGTAT CGACACGTTG
 1051 AACCGCGGTG GCGAATTGTT GATGTGAAAC AGCGCGGGGG TTGTGAAACG
 1101 GACGTGGAAC GCGAACAAACA CGCGAAACAA CGACACGUGG AGCGACAAACA
 1151 ATATTACCTT CGCGCGAACG ATCGAGGAGA GATGTGACAT CGCGCGGGAG
 1201 CGCGCGAAGG CGATGTGACCG CGCGCGGGGGT GAGGGAGGA CGCGCGGGAG
 1251 GAGGAAACATG AGCGCTGTG TGTGAGGGGG CGAGGAGGCG AGCGACCGAC
 1301 ACACCGGCUA CGCGCGAAATG TTGTGAGGGGG CGCGCGGGRA GATCGCGCGAC
 1351 AAGTGGAGAT GTGAGCTGTA GAACTACAAAT GTGTGACCA CGCGCGGGGT
 1401 CGCGCGGGGG CGCGACCGGG CGAGGAGGGGG CGCGCGGGGG CGCGAGAAGC

4/12

1451 CGGCGCCCAT CGGGCGCGTTCCTGGCT TCTTGGGGCG CGGGCGCG
 1501 ACCATGGGGG CGGCCAGGCT GACCTTGACC CTGAGGGCC CGCTGGCTCT
 1551 GAGCGGCATC GTGCAGCAGC AGAACAAACCT CCTGGGGCC ATCGAGGGCC
 1601 AGCAGCATAT GTCAGGCTGACCGCTGTGGG GCACTAAGCA GCTCGAGGCC
 1651 CGCTGGCTCG CGCTGGAGCG CTACCTGAAG GACCGAGGAGC TCTTGGCGCT
 1701 CTGGGGCTCG TCGGCGAAGC TGATCTGACG CACFAACGTA CGCTGGAAACG
 1751 CCTCTGGAG CAACAAGAGC CTGGAGGAGA TCTGGAAACAA CATGACCTGG
 1801 ATCGACTGGG AGCGCGAGAT CGATAACTAC ACCAGCGCTA TCTACAGGCT
 1851 CCTGGAGAG AGCCAGACCG ACCAGGAGAA GAACGAGCAG GAGCTGGCTGG
 1901 ACCTGGACAA CTGGGGAGG CTCTGGAACT CGCTGGACAT CACCAACTGG
 1951 CTCTGGTACA TCAAATCTT CATCATGATT CTGGGGCGC TCTTGGCGCT
 2001 CGCGCTCGTTCGCGCTCG TCTGGCTCGT CGACCGCGT CGGGCGCGCT
 2051 ACAGGGCGCTG GAGCTGGAG ACCGGGGCGCG CGCTGGCGCG CGGGGGCGAC
 2101 CGGGGGCGCG CGATCGAGGA CGAGGGCGCG GAGCGCGACCG CGGACACCG
 2151 CGGGAGGCTG GTGCAGGCTGCT CGCTGGCGAT CGCTGGCGT CGCTGGCGCA
 2201 CGCTGGCTCG GTCAGGCTAC CGACACCGCG ACCGGCTCGT CGGGGGCGCG
 2251 CGGATCTGG AACTCTTAGG CGGGGGCGCG CGGGAGGTGC TGAAGTACTG
 2301 CGGGAAACCTG CGCGAGTATT CGAGCGAGGA CGCTGGCTCG AGCGGGCTGG
 2351 CGCTGGCTCG CGGGGGCGCG ATCGGGCTGG CGGGGGCGAC CGGGGGCGCT
 2401 ATCGAGGCTUC TCGAGAGGGC CGGGGGCGCG ATCGCTGACG CGGGGGCGCG
 2451 CGGGGGCGCG CGGGGGCGCG CGGGGGCGCTCG 3 (SEQ ID NO:39)

FIG. 1
(SHEET 4 OF 4)

5/12

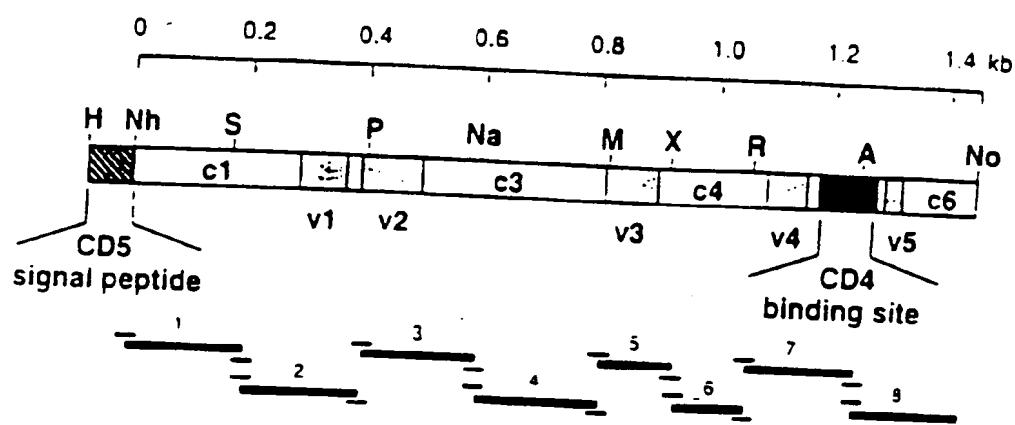


FIGURE 2

6/12

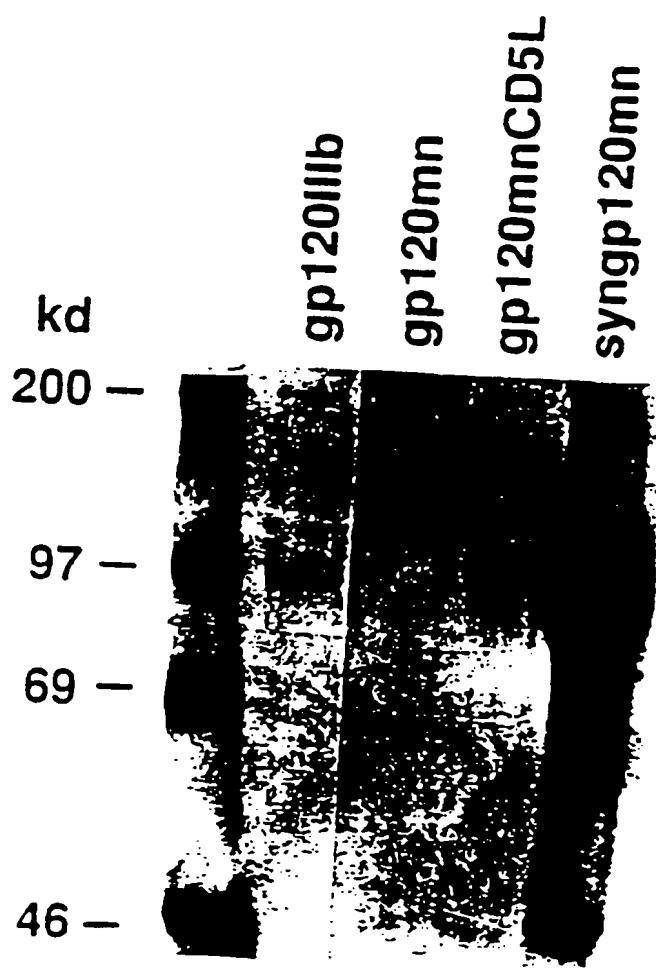


FIGURE 3

7/12

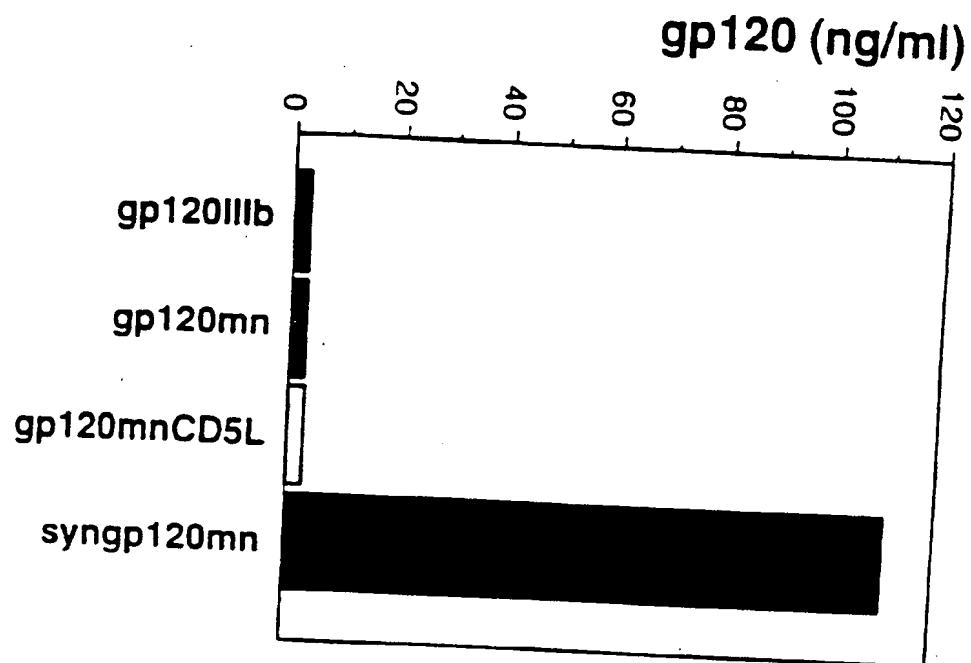


FIGURE 4

8/12

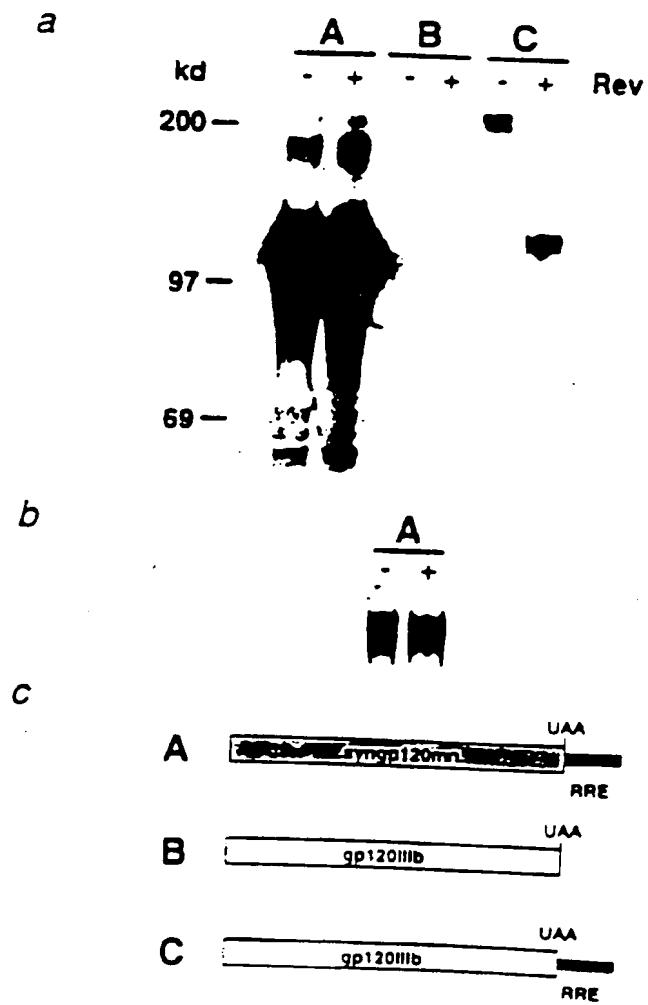


FIGURE 5

9/12

Seq ID# 3C env
 Seq ID# 37-wt
 M N P V I S I T L L S V I Q M S R G Q
 aat cca gta ata agt ata aca tta tta tta agt gta tta caa atg agt aga gga caa
 aac cca gtc alc act ctc ctc tca gtc ctc tca gtc ttt ctt ctt ctt ctt
 K V I S I T A C L V N Q N L R L D C R H
 aga gta ata agt tta aca yca tgt tta gta aat ctt aat ttg aga tta gat tgt aga cat
 aag gtg atc agc ctg aca gtc tgc ctg gtt aa ctt ctt ctt ctt ctt ctt
 E N N T N I P I Q H E F S I T R Q E K K
 gaa aat aca cct ttg cca ata ctt aca cat gaa ttt tca tta acg cgt gaa aaa aaa
 gag aat aac acc ttg ccc atc ctt ctt ctt ctt ctt ctt ctt
 H V L S G T L G V P E H T Y R S R V N L
 cat gta tta agt gga aca tta gga gta cca gaa cat aca ttt aca ttt aca aat
 cac gtg ctg tca ggc acc ctg ggg gtt ccc gag cac act tac ttt ctt ctt
 F S D R F I K V L T I A N F T K D E C
 ttt agt gat aga ttc ata aad gta tta aca ttt gca aat ttt aca aca aat
 ttc agt gac cgc ttt atc aag gtc ctt act ctt gca aac ttt acc acc aag gat gag
 D Y H C E I R V S G Q N P T S S N K T I
 gat tat atg tgt gag ctc agt gta agt gga ctt aat cca aca aca aat
 gac tac atg tgt gaa ctt ctt ctt ctt ctt
 N V I R D K L V K C C C I S I V Q N Y
 aat gta ata aga gat aad tta gta aaa tgt gga aat ctt aca aca aat
 aat gtg atc aga gac aag ctt gtc aag tgt ggt ggc aca aca aca aat
 env agt tgg tta tta tta tta tta agt tta agt ttt ttt ttt ttt ttt
 wt tcc tgg ctg ctg ctc ctc ctc ctc ctc ctc
 S W L I I I I S F I Q A T D F I S
 agt tgg tta tta tta tta tta agt tta agt ttt ttt ttt ttt ttt
 tcc tgg ctg ctg ctc ctc ctc ctc ctc
 env I.
 wt tta tga
 ctg tga

FIGURE 6

10/12

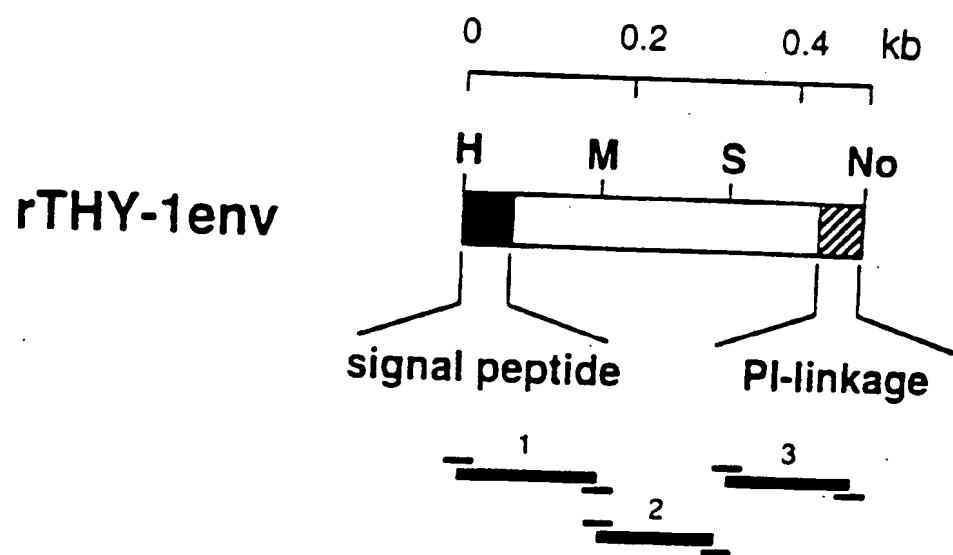


FIGURE 7

11/12

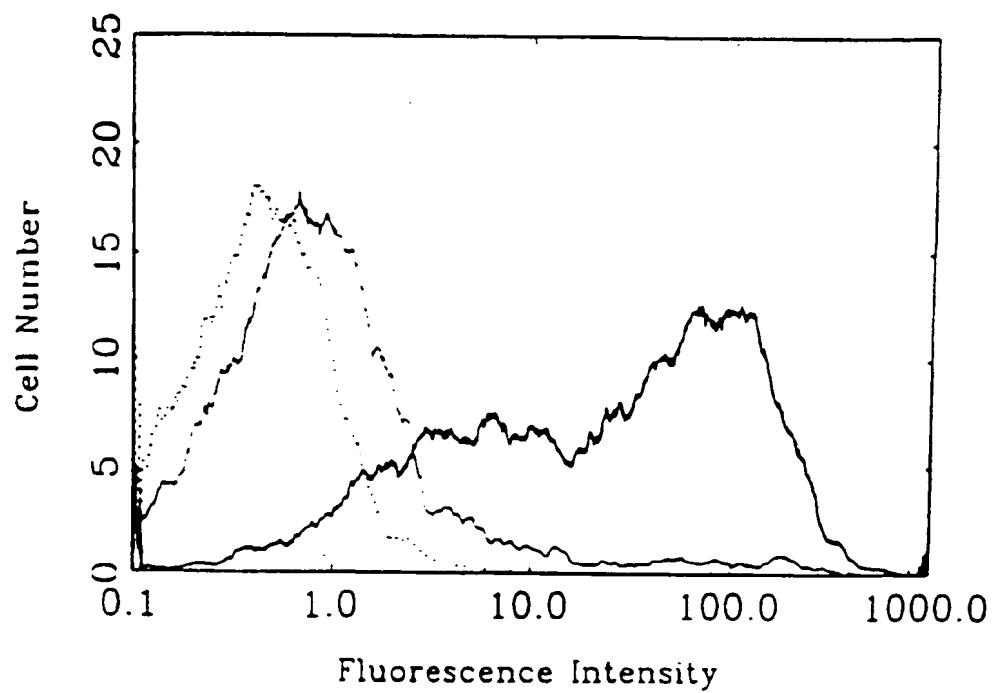


FIGURE 8

12/12

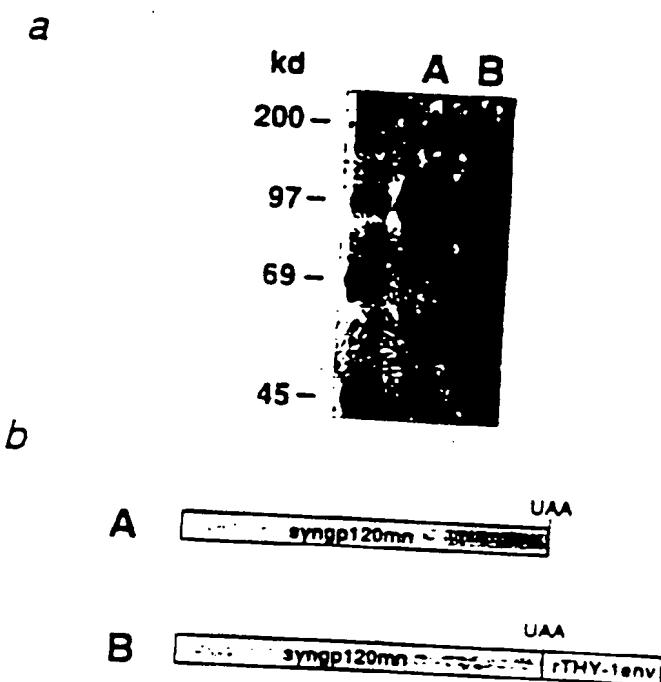


FIGURE 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/11511

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : Please See Extra Sheet.

US CL : Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : Please See Extra Sheet.

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Please See Extra Sheet.

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US, A, 5,270,171 (CERCEK ET AL.) 14 December 1993, see column 34, lines 32-48.	1-16
Y	Nucleic Acids Research, Volume 18, Number 4, issued 1990, McCarrey, "Molecular evolution of the human Pfk-2 retroposon", pages 949-955, see entire document.	1-16
Y	Japanese Journal of Cancer Research, Volume 80, issued March 1989, Kamiya et al., "Transformation of NIH3T3 Cells with Synthetic c-Ha-ras Genes", pages 200-203, see entire document.	1-16

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"T"	
"E" earlier documents published on or after the international filing date	"X"	documents of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"I" document which may throw doubts on priority claim(s) or which is used to establish the publication date of another citation or other special reasons (as specified)	"Y"	document of particular relevance; the claimed invention would be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O" document referring to an oral disclosure, use, exhibition or other means	"A"	document member of the same patent family
"P" document published prior to the international filing date but later than the priority date claimed	"A"	

Date of the actual completion of the international search

24 OCTOBER 1995

Date of mailing of the international search report

03 NOV 1995

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Faximile No. (703) 305-3230

Authorized officer

JAMES KETTER

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

In. national application No.
PCT/US95/11511

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Nucleic Acids Research, Volume 16, Number 17, issued 1988, Sharp et al., "Codon usage patterns in <i>Escherichia coli</i> , <i>Bacillus subtilis</i> , <i>Saccharomyces cerevisiae</i> , <i>Schizosaccharomyces pombe</i> , <i>Drosophila melanogaster</i> and <i>Homo sapiens</i> : a review of the considerable within-species diversity", pages 8207-8211, see entire document.	1-16
Y	Proceedings of the National Academy of Sciences USA, Volume 83, issued November 1986, Newgard et al., "Sequence analysis of the cDNA encoding human liver glycogen phosphorylase reveals tissue-specific codon usage", pages 8132-8136, see entire document.	1-16
Y	Gene, Volume 46, issued 1986, Coulombe et al., "Expression of a synthetic human interferon- α , gene with modified nucleotide sequence in mammalian cells", pages 89-95, see entire document.	1-16

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/11511

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

C12N 15/09, 15/12, 15/33, 15/64

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

536/23.5, 23.72; 435/172.3

B. FIELDS SEARCHED

Minimum documentation searched

Classification System: U.S.

536/23.5, 23.72; 435/172.3

B. FIELDS SEARCHED

Documentation other than minimum documentation that are included in the fields searched:

NONE

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, MEDLINE EXPRESS